FUSING MULTIPLE HEART RATE SIGNALS TO REDUCE ALARMS IN THE ADULT INTENSIVE CARE UNIT

by

Kasey Benson Poon

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

in

Medical Informatics/Health Services Administration

Department of Medical Informatics

The University of Utah

May 2005

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of the University of Utah

I have read the thesis of Benson Poon in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

Date

Reed M. Gardner
Chair: Supervisory Committee

Approved for the Major Department

Reed M. Gardner
Chair/Dean

David S. Chapman
Dean of the Graduate School

ABSTRACT

The objective of the study was to compare the performance of two physiologic monitor algorithms in their ability to generate true heart rate alarms and to avoid producing false alarms. The "standard" algorithm, the algorithm currently used in GE/Marquette monitors installed at LDS Hospital, was compared with a "fusion" algorithm that combined heart rate data independently determined from electrocardiogram (ECG), intra-arterial pressure (ART), and pulse oximeter signals.

Data were collected from patients admitted to the medical, surgical, and cardiac ICUs at LDS Hospital in Salt Lake City, Utah, from April through September, 2001. Data from a total of 109 patients were collected for periods of up to 24 hours for each patient, which resulted in a total of 1902.25 patient-hours of data. The physiological signals were then presented to both algorithms to allow a direct comparison of both alarm methods. A physician reviewed the heart rate alarm results of each algorithm and determined whether each alarm generated was true or false. Since alarm conditions were only studied if they were detected by one of the algorithms, we were not aware of any condition that should have generated an alarm but did not. The following five alarm conditions were studied: low heart rate (LHR), high heart rate (HHR), asystole, ventricular tachycardia (VT), and ventricular fibrillation (VF).

The "standard" algorithm generated 341 alarms; 118 (34.6%) were true and 223 (65.4%) were false. The fusion algorithm produced 184 alarms; 126 (68.5%) were true

and 58 (31.5%) were false. There were 149 instances in which both algorithms produced the same alarms. Of these 149 instances 111 (74.5%) were true alarms and 38 (25.5%) were false. Of the combined total of 525 alarms (341 standard + 184 fusion), 316 (60.2%) had durations of 10 seconds or less.

To determine if a patient's average heart rate was associated with low or high heart rate alarms, the average heart rate was calculated. Average heart rate was determined by summing all heart rate values and dividing by the total number of heart rates for each patient's data set. Of the 267 LHR alarms, 148 (55.4%) were from patients who had average heart rates of 80 beats per minute (BPM) or less. Of the 127 HHR alarms, 114 (89.8%) were from patients who had average heart rates of 90 BPM or greater.

While there was no "gold standard" test available to calculate the actual sensitivity and false positive rate of each algorithm, we were able to compare the two algorithms to each other using relative sensitivity (RSN) and relative false positive rate (RFP). The RSN of the fusion algorithm compared to the standard algorithm was 1.09 (95% CI 1.01 to 1.17). The 95% confidence interval (CI) for the RSN indicated that there was a 95% chance that the true positive rate for the fusion algorithm was between 1% and 17% greater than the standard algorithm. The RFP of the fusion algorithm compared to the standard algorithm was 0.27 (95% CI 0.21 to 0.34). The 95% CI for the RFP indicated that there was a 95% chance that the false positive rate for the fusion algorithm was between 66% and 79% lower than the standard algorithm.

Using the fusion algorithm to reduce false alarms did not result in a concomitant increase in the number of true alarms that were "missed" by the fusion algorithm. On the contrary, the fusion algorithm missed only 5 true alarms that the standard algorithm caught, while the standard algorithm missed 15 true alarms that the fusion algorithm caught.

Combining redundant physiologic signal measurements through the use of a fusion algorithm was an effective way to eliminate false positive alarms. Because the fusion algorithm was more sensitive, had a higher positive predictive value, and missed fewer true alarms than the standard algorithm, it was determined to be a superior method of generating alarms. In addition, we found that simply delaying alarms by 10 seconds could result in a 60.2% reduction in false alarms. Considering the patient's average heart rate when setting low or high heart rate alarm thresholds could also reduce false alarms. All of these methods should be integrated into new bedside monitors to reduce false alarms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Excessive numbers of physiological monitoring alarms, many of which are false, in intensive care units (ICUs) are nothing new. In a study published over three decades ago, Raison *et al.* noted that ". . . occurrence of false alarms from automated patient monitor systems is a common clinical problem."[1] More recent studies and reviews have validated these findings.[2-4] Not surprisingly, a large number of alarms in the ICU are either false or are clinically irrelevant.[5-7] The large number of false alarms produced by physiological monitors likely results in many true alarms being ignored.

The danger with the generation of so many false alarms is that clinicians may become desensitized to them or irritated by them. Frustration over false alarms may even lead physicians and nurses to deactivate the alarms, effectively rendering the physiologic monitors useless as alerting devices. For instance, surveys have revealed that many anesthesiologists turn off alarms at the beginning of surgical cases. The leading reason for this behavior is the feeling that the devices generate too many false alarms.[8-9] Other reasons include confusion caused by multiple monitors generating different alarm tones, the need for a quiet work environment, and the unpleasant sound of the alarms.[8]

Part of the problem of excessive alarms is that numerous devices may be monitoring each patient. Each machine operates independently of the others, and often multiple devices are measuring the same physiologic parameters. As noted by Duberman and Bendixen, "it is . . . true that the greater the number of detection devices added to the

system, the greater the likelihood that one of these devices will yield a false-positive reading."[10]

Efforts have been made to reduce the excessive number of false alarms in the intensive care unit. Raison, *et al.* proposed a "combined" system in which a heart rate alarm would be triggered only when limits of both an electrocardiogram (ECG) device and an arterial pulse monitor were exceeded.[1] Feldman, Ebrahim, *et al.* suggested a similar strategy to reduce false positives by combining multiple, independent heart rate signals to create a single, "fused" signal. [11-12] The algorithm used to produce the fused signal was termed "Robust Sensor Fusion" (RSF). The results of the RSF study indicated ". . . that fused estimates of heart rate obtained using RSF are consistently better than the estimates that can be obtained from any individual sensor. Not only is the quality of the heart rate estimates improved, but fewer false high and low heart rate alarms will occur if the alarms are based upon the fused estimate of heart rate."[11-12]

In the present study a similar fusion algorithm was used to analyze patients' heart rates and rhythms. The performance of the fusion algorithm was compared to that of the standard algorithm used by the GE/Marquette monitors in our intensive care units.

MATERIALS AND METHODS

This study was conducted in collaboration with GE/Marquette, a major manufacturer of bedside ICU physiologic monitors. GE/Marquette provided both funding for the study and the fusion algorithm that was used for comparison to the standard algorithm that is currently being used in the monitors. The fusion algorithm was proprietary to GE/Marquette. Therefore, the investigators in the study were not privileged to know the methods used to combine multiple physiological signals to determine a single heart rate and to generate an alarm. Three physiological monitors were used to detect heart rate alarms in our study:

Electrocardiogram (ECG)
Intra-arterial pressure (ART)
Pulse Oximeter

Data were collected from patients admitted to the thoracic surgery, shock/trauma, hyperbaric, medical/surgical, and cardiac ICUs at LDS Hospital. Located in Salt Lake City, Utah, LDS Hospital is a 467-bed, nonprofit, tertiary care center. The data used in this study were collected as part of a "false alarm" quality-improvement project at the hospital. The monitors used at LDS Hospital are GE/Marquette Solar 8000 models. The study was conducted over a 6-month period from April through September, 2001. Data collection was entirely passive; that is, the process involved simply attaching monitoring devices to patients and recording their physiologic information onto a hard disk drive without any action required on the part of the patient or clinician.

The Institutional Review Boards (IRBs) of both LDS Hospital and the University of Utah approved the study following independent, expedited reviews of the study application. Since data collection required no change in patient therapy and no protected health information (PHI) was collected, we applied for a waiver of the requirements to obtain patient consent. Each IRB independently granted the waiver.

Five alarm conditions were examined during the study: low heart rate (LHR), high heart rate (HHR), asystole, ventricular tachycardia (VT), and ventricular fibrillation (VF). Other alarms, such as those caused by other arrhythmias or by lead failures, were not evaluated.

Data were collected from a total of 109 ICU patients, as noted in Table 1, who had the following sensors attached:

ECG Monitor
Intra-arterial pressure (ART)
Pulse Oximeter

All patients who met the above criteria and who were admitted to the ICUs at LDS Hospital during the study period of April through September, 2001, were included in the study. The disparity between the number of patients from the thoracic surgery and shock

Table 1. Number of patients and total patient-hours for data collection from each intensive care unit.

Unit	Thoracic Surgery	Shock Trauma	Hyperbaric	Medical/ Surgical	Cardiac	**Total**
Number of Patients	71	31	4	2	1	**109**
Total Patient-Hours	562.11	1304.16	6.98	5.00	24	**1902.25**

trauma units compared to hyperbaric, medical/surgical, and cardiac units reflected the difference in the number of patients who had intra-arterial catheters inserted. The ECG, ART, and pulse oximeter all detected and displayed patients' heart rates. Additionally, the ECG monitored patients' heart rhythm, the ART monitored patient's blood pressure, and the pulse oximeter monitored patients' blood oxygen saturation. Data were recorded on each patient for up to 24 hours, and the study group covered 1902.25 total patient-hours. The derived heart rate data was recorded every 2 seconds, resulting in 3,424,050 heart rate data points. The standard alarm algorithm and the fusion algorithm were both applied to exactly the same physiologic data. The experiment allowed direct comparison of the two algorithms' performance given that they were presented with exactly the same physiologic data. Data analysis was conducted offline, not in real time.

A personal computer running the fusion software was connected to the Marquette Unity network. All of the hospital's patient monitors and central monitoring stations were attached to the same network. Physiologic data were recorded onto the local hard drive of the personal computer that contained the fusion algorithm. Data that were recorded consisted of the following:

ECG data from leads I, II, III, and V monitoring heart rate and rhythm signals
Arterial blood pressure data
Pulse oximeter data
Standard alarm algorithm data
Fusion alarm algorithm data

Figure 1 shows a sample screenshot of the physiologic waveforms that were recorded from a patient (ie ECG leads, ART, and pulse oximeter).

The standard and fusion alarm algorithm data included the patient's heart rate; alarm start and end times recorded in hours, minutes, and seconds; and alarm condition

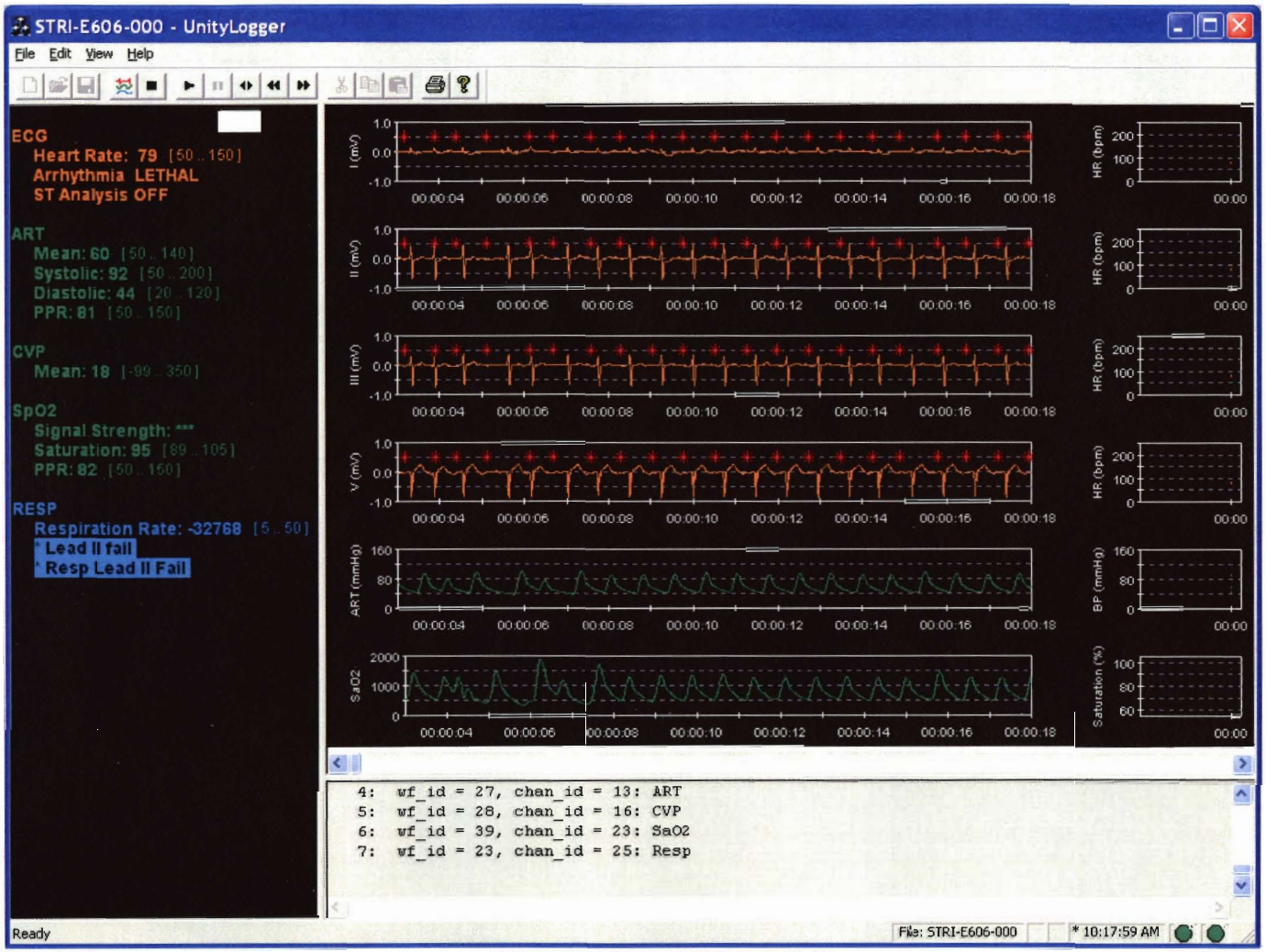

Figure 1. Screenshot of sample physiologic waveforms recorded from a patient. The following waveforms are shown: Leads I, II, III, and V of the ECG; intra-arterial pressure (ART); and pulse oximeter (SaO2).

(LHR, HHR, asystole, VT, and VF). The fusion algorithm combined data from all three physiological signals and at times produced a heart rate different from the standard algorithm, which was ECG-based. While it was possible that caregivers may have silenced the actual alarm tones, the alarm conditions that originally generated the tones were still logged for as long as they existed. Therefore, the total duration of the alarm conditions were recorded, not simply the duration of the alarm tone.

Figure 2 shows a schematic of the structure of the data collection network. The monitors' default settings for low and high heart rate alarms were 50 beats per minute (BPM) and 150 BPM. These thresholds could be manually changed by caregivers. If an alarm condition occurred and ended before any intervention by a caregiver, the alarm tone would cease, but the monitor display screen would continue to indicate that an alarm had been triggered. Thus, the display provided a memory of alarms that had occurred even if a clinician had not been at the bedside when the alarm occurred. Data were recorded from each patient for a maximum of 24 hours or for a shorter time interval if the patient expired or was transferred to an unmonitored bed.

Figure 3 shows a flow chart of the steps taken in the evaluation process. Each alarm was reviewed by a physician trained in internal medicine (Dr. Poon) to determine whether the alarm was a true positive or a false positive. This physician review served as the reference standard in distinguishing true alarms from false alarms. If one of the heart rate algorithms alarmed, and the other did not, then the algorithm that did not alarm was reviewed to determine if the non-alarming algorithm was a true negative or false negative. During times when neither a standard nor a fusion alarm was activated, the data were not evaluated to determine whether the lack of alarms was due to true negative

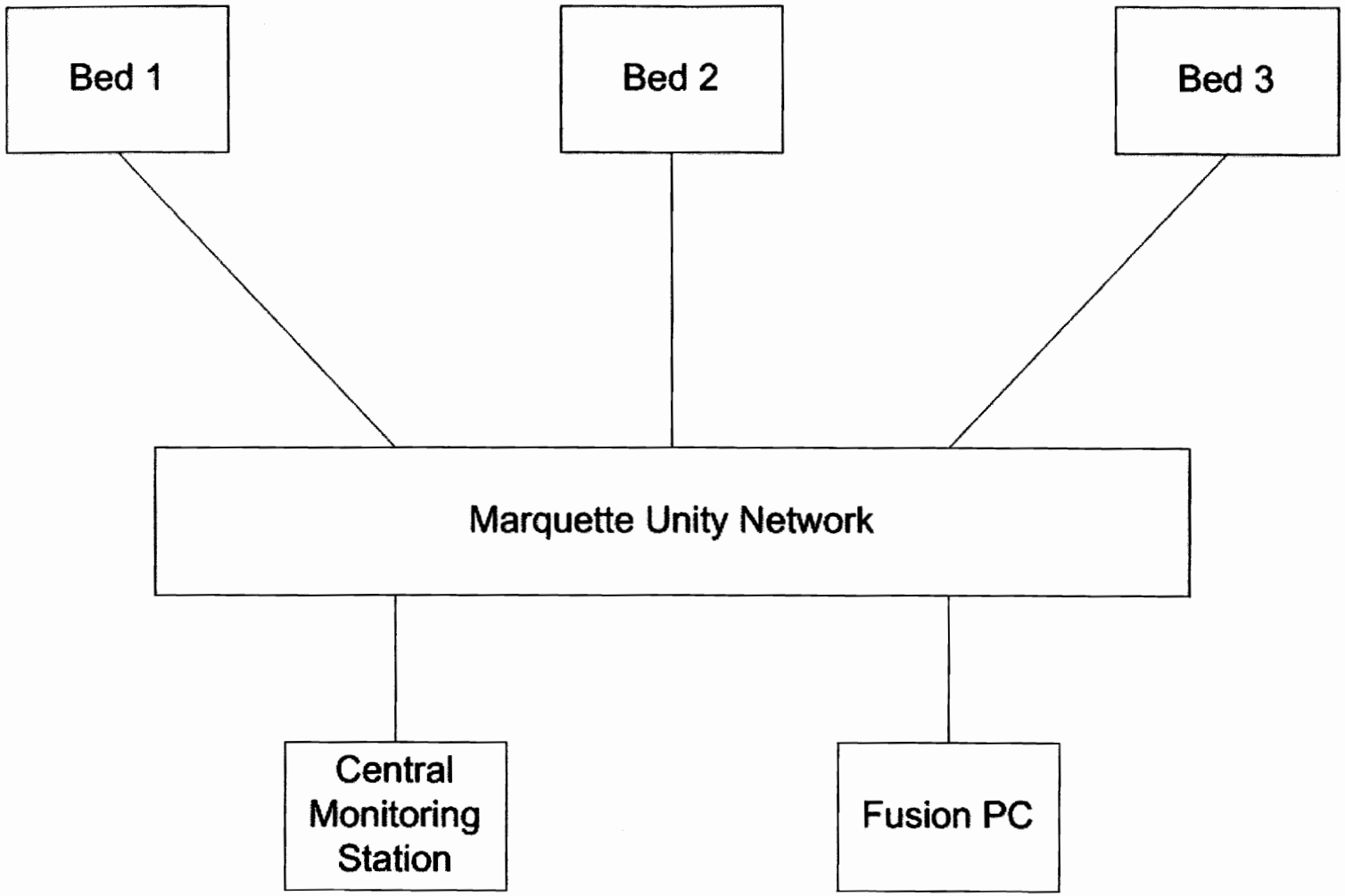

Figure 2. Structure of the monitoring network used for data collection. The Central Monitoring Station monitors applied only the standard alarm algorithm, while the Fusion PC applied both the standard and the fusion algorithms and recorded results of both algorithms on its hard drive.

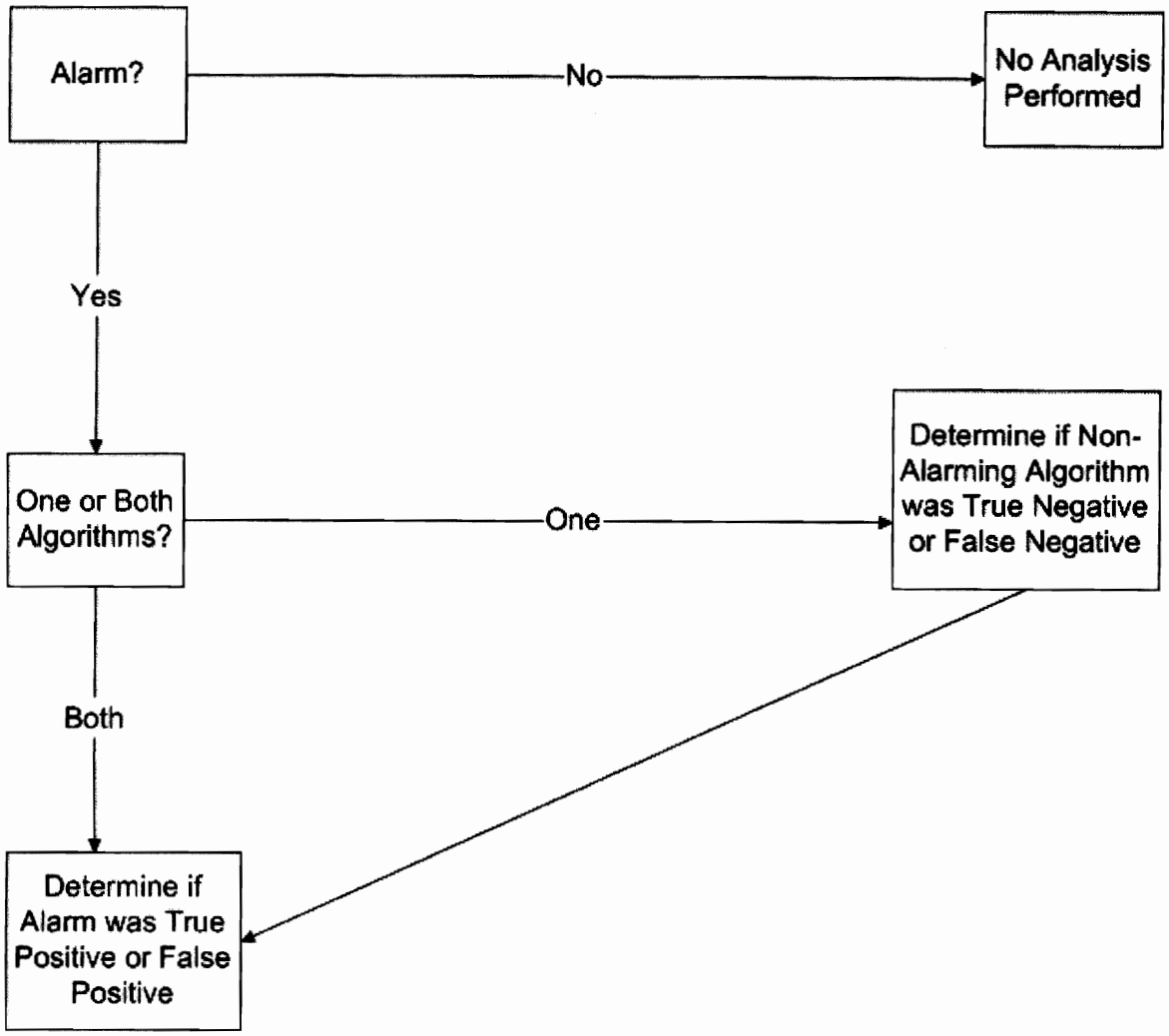

Figure 3. Data analysis process. If neither algorithm alarmed, there was no analysis performed to determine true negative or false negative status. Only when one or both algorithms alarmed were analyses done to determine true positive, false positive, true negative, or false negative status.

or false negative evaluation by the algorithms. Physician determination of an algorithm's true negative or false negative status was only performed in instances in which one algorithm alarmed while the other did not. The reason for evaluating the algorithms in this manner was the lack of a true "gold standard" control against which the two algorithms' outputs could be compared.

RESULTS

A standard alarm was one which was triggered by the standard algorithm currently used by the GE/Marquette physiologic patient monitors. The algorithm relied on data from a single monitoring signal, typically the patient's ECG waveforms. A fusion alarm, on the other hand, was an alarm triggered using the fusion algorithm. The fusion algorithm processed signals from three different monitoring signals--ECG, intra-arterial pressure (ART), and pulse oximeter--to produce a single, "fused" heart rate.

Of the 109 patients in the study sample, 51 triggered one or more of the five alarm types recorded. There were 376 instances in which alarms were triggered. Of these 376 instances 192 were only standard alarms, 35 were only fusion alarms, and 149 were both standard and fusion alarms (Figure 4). Of the 149 situations in which both algorithms

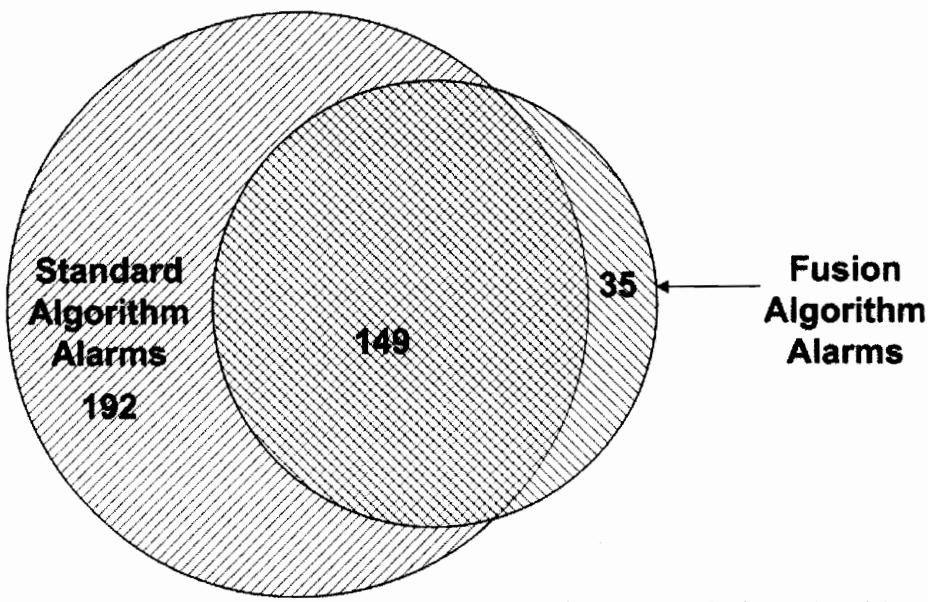

Figure 4. Instances in which the standard algorithm, the fusion algorithm, or both algorithms alarmed (n=376).

generated the same alarms, 111 were true alarms and 38 were false. Summing all the standard alarms (192 + 149) and all the fusion alarms (35 + 149) resulted in a total of 525 individual alarms. Of the 525 total individual alarms 244 were true positive alarms and 281 false positives.

Table 2 shows the distribution of true positive and false positive alarms from the standard and fusion algorithms. Almost two-thirds of all the standard alarms were false (65.4%), while only roughly one-third of the fusion alarms were false (31.5%). The positive predictive value (PPV)--calculated using the true positive and false positive numbers in the table--of the standard algorithm was 34.6%. The PPV of the fusion algorithm was 68.5%. The PPV of 34.6% for the standard algorithm is similar to the PPV of 27% which Chambrin *et al.* found in their study.[5]

Table 3 shows the amount of time occupied by alarms generated by the standard and fusion algorithms. The standard algorithm generated 2.62 hours of alarms, and 83.2% of those were false positives. The fusion algorithm, though, only generated 1 hour of alarms, and only 54.0% of those were false positives.

Table 2. Total number of true positive (TP) and false positive (FP) alarms for standard and fusion algorithms.

	Standard Alarms (% of Total)	**Fusion Alarms (% of Total)**	**Total Alarms**
True Positive (TP)	118 (34.6%)	126 (68.5%)	244
False Positive (FP)	223 (65.4%)	58 (31.5%)	281
Total	341 (100%)	184 (100%)	525

Table 3. Total time in hours spent alarming under both algorithms for true positive (TP) and false positive (FP) alarms.

	Standard Alarms Time (% of Total)	Fusion Alarms Time (% of Total)	Total Alarms Time (hours)
True Positive (TP)	0.44 (16.8%)	0.46 (46.0%)	0.90
False Positive (FP)	2.18 (83.2%)	0.54 (54.0%)	2.72
Total (hours)	2.62 (100%)	1.00 (100%)	3.62

The details of the 341 standard and 184 fusion alarms recorded from the 51 patients who triggered alarms are shown in Table 4 and Table 5. The tables show that the greatest numbers of alarms that were generated in this study were for Low HR [n = 262 (192+70)] and High HR [n = 128 (64+64)]. The ratio of standard algorithm alarms to fusion algorithm alarms for High HR, VT, and VF were 64:64, 26:21, and 5:4, respectively. Tables 4 and 5 also show that for High HR, VT, and VF, the ratios of true positive to false positive alarms was similar for both algorithms. However, the standard algorithm produced 29 (54-25) more asystole and 122 (125-70) more Low HR alarms than did the fusion algorithm. Furthermore, 43% of all standard alarms were false positive Low HR alarms, and 16% of all standard alarms were false positive asystole

Table 4. Number of true positive (TP) and false positive (FP) alarms using the standard algorithm for each of five alarm conditions studied (HR=heart rate, VT=ventricular tachycardia, VF=ventricular fibrillation). The bottom row of percentages shows the percentage out of the total number of alarms that the given alarm represents.

341 Standard Alarms from 51 Patients												
	Low HR		**High HR**		**Asystole**		**VT**		**VF**		**Total**	
	192		64		54		26		5		341	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
n	45	147	50	14	1	53	18	8	4	1	118	223
% of Total	13%	43%	15%	4%	<1%	16%	5%	2%	1%	<1%	35%	65%

Table 5. Number of true positive (TP) and false positive (FP) alarms using the fusion algorithm for each of five alarm conditions studied (HR=heart rate, VT=ventricular tachycardia, VF=ventricular fibrillation). The bottom row of percentages shows the percentage out of the total number of alarms that the given alarm represents.

184 Fusion Alarms from 51 Patients												
	Low HR		**High HR**		**Asystole**		**VT**		**VF**		**Total**	
	70		64		25		21		4		184	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
n	48	22	55	9	1	24	18	3	4	0	126	58
% of Total	26%	12%	30%	5%	<1%	13%	10%	2%	2%	0%	68%	32%

alarms. On the other hand, only 12% of all fusion alarms were false positive Low HR alarms, and only 13% of all fusion alarms were false positive asystole alarms.

During data collection, if a nurse or other caregiver turned off or reset a monitor's audible alarm tone, the monitor continued to record the alarm condition for as long as it existed. Therefore, the timestamps on the alarms reflected the true durations of the physiologic alarm conditions, not simply the durations of the audible alarm tones.

During data analysis, it was noted that many alarms lasted for only 10 seconds or less. Therefore, the duration in seconds of each of the 525 total alarms was determined, and the numbers of alarms in each time increment were counted. Figure 5 through Figure 7 show the number of alarms for each time increment.

The greatest number of alarms lasted 10 seconds or less. For all alarm conditions 60.2% were for 10 seconds or less; for standard alarm conditions 55.1% were for 10 seconds or less; and for fusion alarm conditions 69.6% were for 10 seconds or less. Furthermore, review of Figures 5, 6, and 7 show that there was a dramatic drop-off of alarms of longer duration. The drop-off seen was independent of whether the alarms were true or false and whether they were generated by the standard or fusion algorithms.

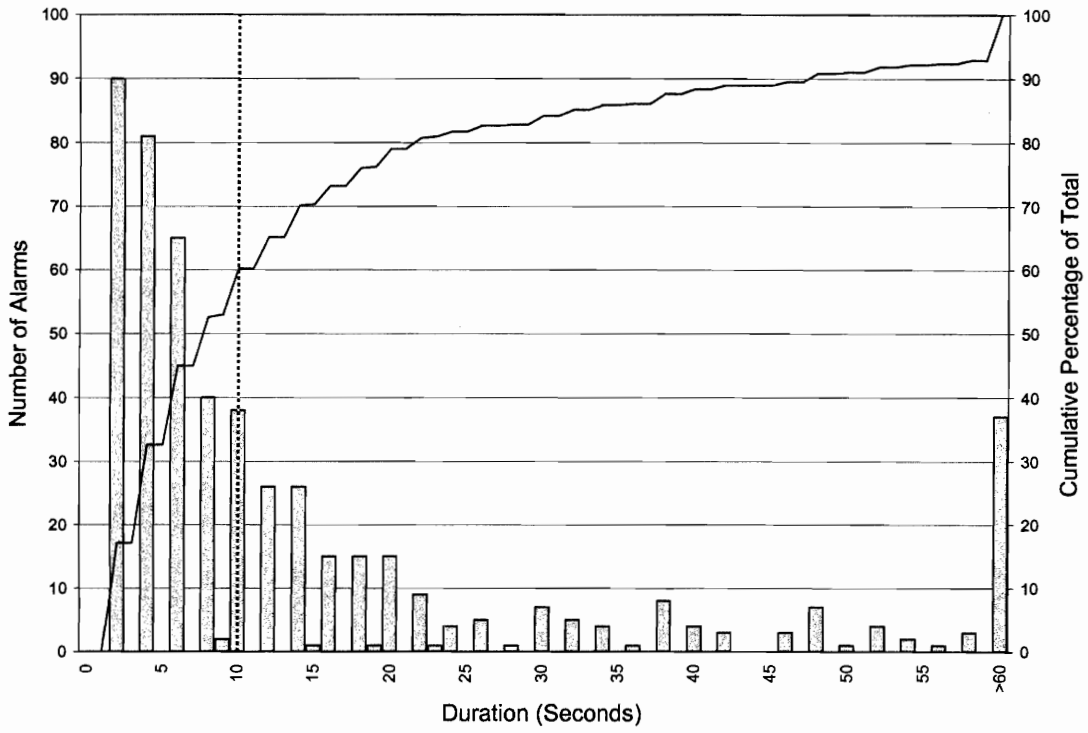

Figure 5. Total number of all alarms by duration (n=525). The dotted vertical line indicates the 10-second duration.

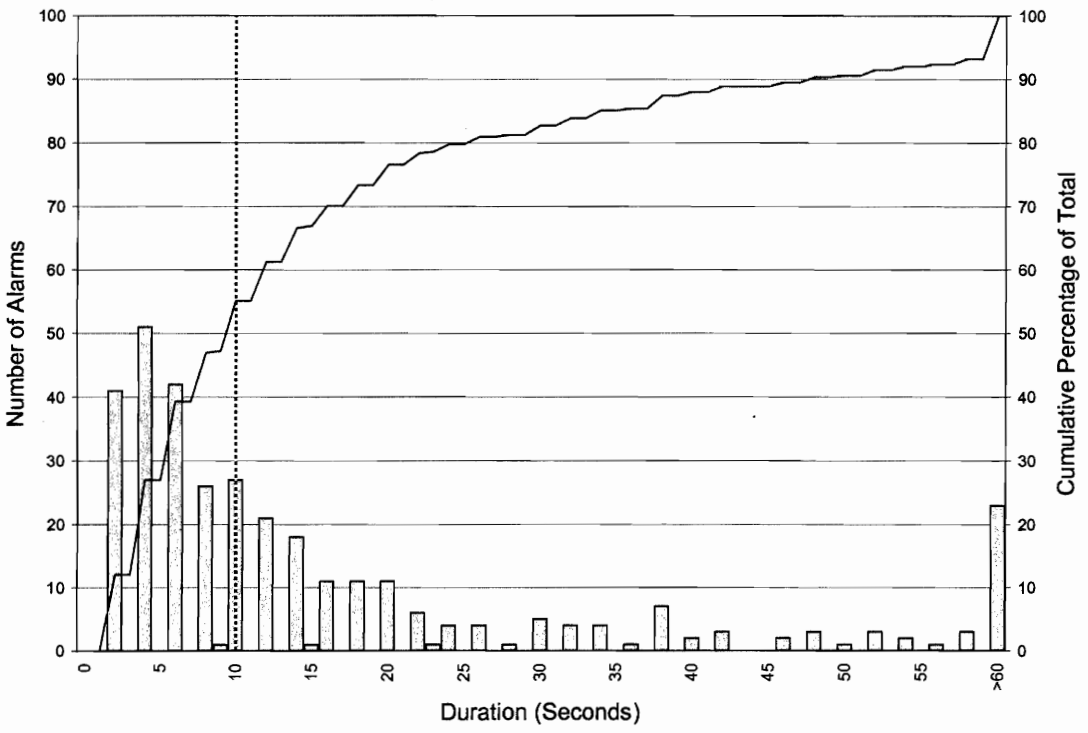

Figure 6. Total number of standard alarms by duration (n=341). The dotted vertical line indicates the 10-second duration.

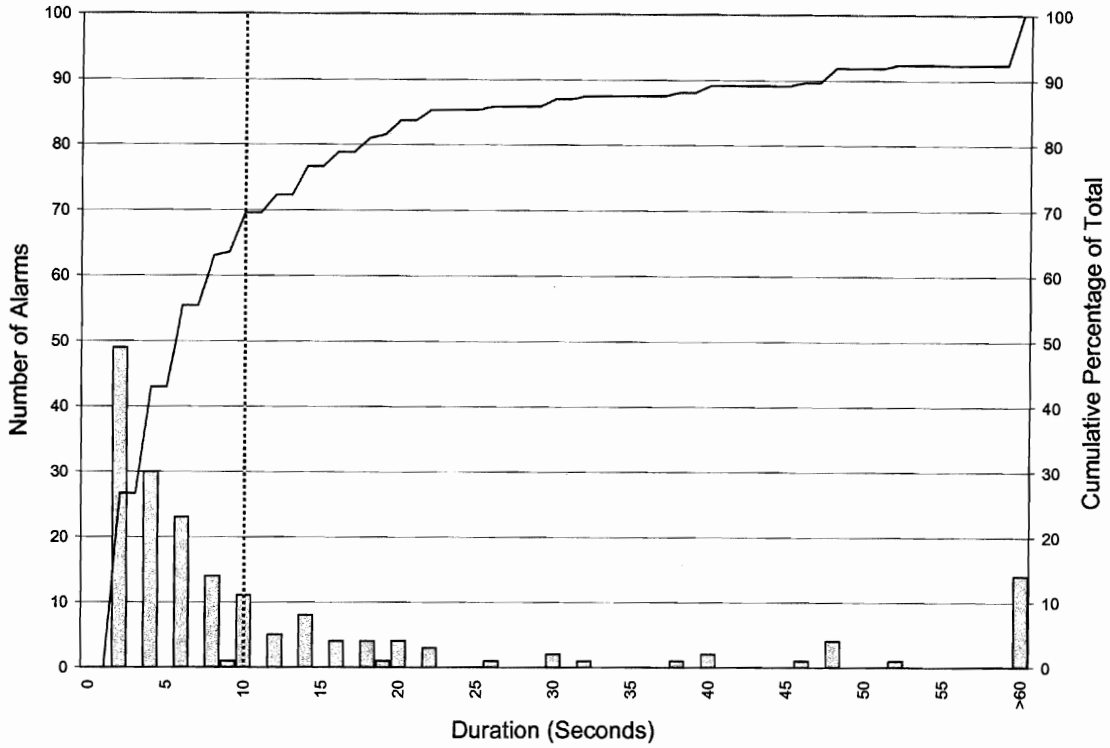

Figure 7. Total number of fusion alarms by duration (n=184). The dotted vertical line indicates the 10-second duration.

There were more standard algorithm alarms (n=188) of 10 seconds or less than there were fusion algorithm alarms (n=128) of 10 seconds or less.

To determine if a patient's average heart rate had any association with the number of low and high heart rate alarms generated by the patient, the average heart rates of all patients who generated alarms were calculated. The average heart rate was calculated by dividing the sum of all heart rates recorded for a patient by the total number of heart rate values recorded for that patient. The total numbers of true positive alarms as functions of average heart rates were then plotted. These results are shown in Figure 8 and Figure 9. Figure 8 illustrates that patients who had an average heart rate between 60 and 80 beats per minute had a higher number of true positive low heart rate alarms than did patients with higher average heart rates. Figure 9, on the other hand, shows that patients who had

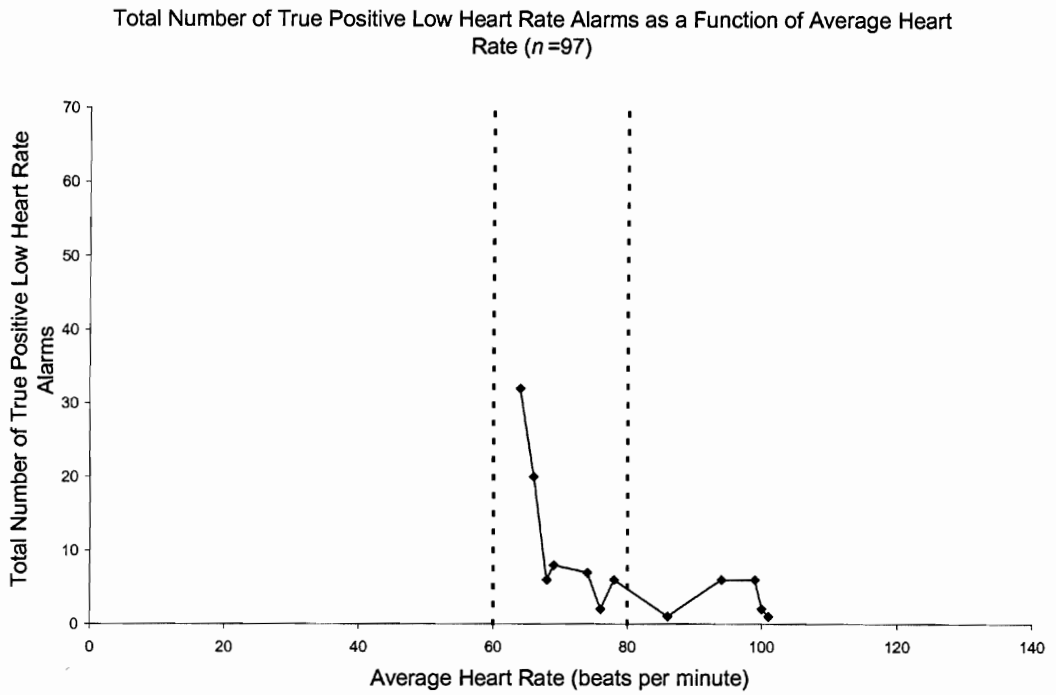

Figure 8. The greatest number of true positive low heart rate alarms occurred in patients with average heart rates in the range, indicated by the dotted vertical lines, between 60 and 80 beats per minute.

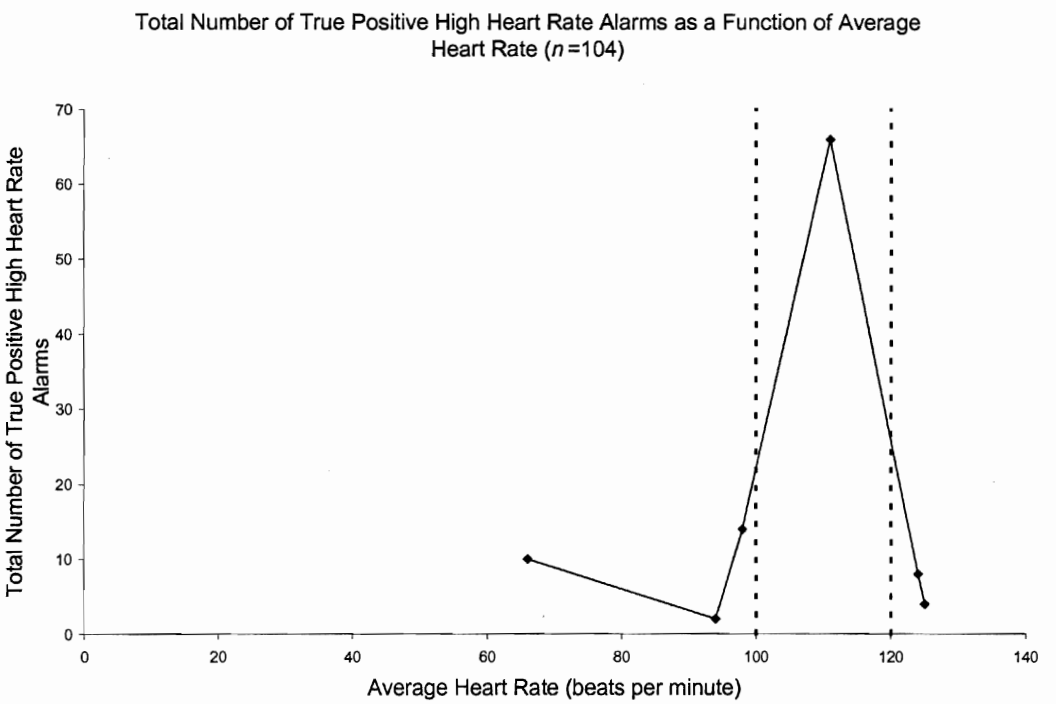

Figure 9. The greatest number of true positive high heart rate alarms occurred in patients with average heart rates in the range, indicated by the dotted vertical lines, between 100 and 120 beats per minute.

an average heart rate between 100 and 120 beats per minute had a higher number of true positive high heart rate alarms than did patients with lower average heart rates.

The default alarm limits of the monitors for low and high heart rates were 50 BPM and 150 BPM, respectively. Of the 109 patients in the study, 51 patients triggered alarms and 58 did not. Of the 58 patients who did not trigger alarms, 50 were patients for whom the alarm limits were never changed from the low of 50 and the high of 150 beats per minute. Of the 8 patients whose monitor alarm limits were changed from their defaults, 3 had only the low limit changed, and 5 had only the high limit changed. Of the 51 patients who did trigger alarms, there were 39 patients in whom the alarm limits were never changed from their default settings. Of the 12 patients in whom monitor alarm limits were changed from their defaults, 3 had only the low limit changed, 7 had only the high limit changed, and 2 had both changed. Table 6 summarizes these results. Of the 109 total patients, 89 (81.7%) never had their monitors' alarm limits changed from their default values.

Table 6. Number and percentage of changes in heart monitor alarm limit defaults for 109 adult intensive care unit patients.

Alarm Limit Defaults	Alarm	No Alarm
No Change	39 (76.5%)	50 (86.2%)
Low Limit Only	3 (5.9%)	3 (5.2%)
High Limit Only	7 (13.7%)	5 (8.6%)
Both Low and High Limits	2 (3.9%)	0 (0%)
Total	51 (100%)	58 (100%)

Relative Sensitivity and Relative False Positive Rate

The two algorithms were further compared by calculating relative sensitivity (RSN) and relative false positive rate (RFP). RSN and RFP allow for meaningful comparisons of two tests when a true reference standard cannot be applied. Calculation of RSN and RFP has been previously described by others.[13-14] Table 7 shows contingency tables that can be used to organize data for the calculation of RSN and RFP. In Table 7 the hypothetical sum a+b+c+d is the total number of situations where alarm conditions are truly present, a'+b'+c'+d', or truly absent, a"+b"+c"+d" (a=a'+a", b=b'+b", c=c'+c", and d=d'+d"). Alarm conditions can be truly present or truly absent regardless of whether or not a physiologic monitor detects them. The total number of instances where alarm conditions are truly present is represented by the sum a'+b'+c'+d', which also equals all true positive (TP) results plus all false negative (FN) results. The total number of instances where alarm conditions are truly absent is represented by the sum a"+b"+c"+d", which also equals all false positive (FP) results plus all true negative (TN) results. Because a true "gold standard" does not exist for detecting the presence or absence of alarm conditions, the values d' and d" cannot be known. Therefore, the sums a'+b'+c'+d' and a"+b"+c"+d" cannot be known either. The sum a+b+c+d, being composed of two unknown values--a'+b'+c'+d' and a"+b"+c"+d"--is, therefore, also unknown.

Furthermore, without knowing the value of a'+b'+c'+d' (ie TP+FN), it is impossible to determine sensitivity [ie TP/(TP+FN)]. Similarly, without knowing the value of a"+b"+c"+d" (ie TN+FP), it is impossible to determine specificity [ie TN/(TN+FP)]. However, using physician confirmation as a reference standard, the true

Table 7. Data layout of initial and confirmed alarm conditions using standard and fusion algorithms. Variables in brackets represent unknown values (adapted from Cheng and Macaluso).[13]

A. Initial Results of Two Algorithms

Fusion	Standard +	Standard -	
+	a	b	a+b
-	c	d	c+d
	a+c	b+d	a+b+c+d

B. Alarm Conditions Truly Present

Fusion	Standard +	Standard -	
+	a'	b'	a'+b'
-	c'	[d']	[c'+d']
	a'+c'	[b'+d']	[a'+b'+c'+d']

C. Alarm Conditions Truly Absent

Fusion	Standard +	Standard -	
+	a"	b"	a"+b"
-	c"	[d"]	[c"+d"]
	a"+c"	[b"+d"]	[a"+b"+c"+d"]

status of only those patients who actually generate an alarm can be determined. The confirmatory procedure results in a reclassification of alarm patients into the categories listed in Tables 7B and 7C. Among the patients who generated true positive alarms, RSN can be calculated using the following formula:

$$RSN = \frac{(a'+b')/(a'+b'+c'+d')}{(a'+c')/(a'+b'+c'+d')} = \frac{a'+b'}{a'+c'}.$$

Among patients confirmed to be truly absent of alarms conditions, RFP can be calculated with the following equation:

$$RFP = \frac{(a''+b'')/(a''+b''+c''+d'')}{(a''+c'')/(a''+b''+c''+d'')} = \frac{a''+b''}{a''+c''}.$$

The RFP is associated with the specificity of each test by the following relation:

$$RFP = \frac{1-F_1}{1-F_2}.$$

Table 8 shows the alarms generated by the standard and fusion algorithms separated into groups where alarms were truly present or truly absent. The variance of RFP can also be determined by replacing a', b', and c' with a", b", and c", respectively. Once the variance is known, confidence intervals can be calculated. Using the values in Table 8, the RSN of the fusion algorithm to the standard algorithm was calculated to be 126/116 = 1.09. Therefore, the fusion algorithm is 9% more sensitive than the standard algorithm. The 95% confidence interval for the RSN was 1.01 – 1.17. Similarly, the RFP was determined to be 59/222 = 0.27, so the fusion algorithm resulted in 73% fewer

Table 8. Results of standard and fusion algorithms by true alarm status.

A. Alarm Conditions Truly Present

		Standard +	Standard -	
Fusion	+	111	15	126
	-	5	[d']	[c'+d']
		116	[b'+d']	[a'+b'+c'+d']

B. Alarm Conditions Truly Absent

		Standard +	Standard -	
Fusion	+	38	21	59
	-	184	[d"]	[c"+d"]
		222	[b"+d"]	[a"+b"+c"+d"]

false positives than the standard algorithm. The 95% confidence interval for the RFP was 0.21 – 0.34.

Finally, Table 8A shows that the fusion algorithm failed to alarm in only 5 instances where the standard algorithm generated a true positive alarm. Conversely, the standard alarm algorithm failed to generate an alarm in 15 situations where a true positive alarm was generated by the fusion alarm algorithm.

DISCUSSION

The results of this study indicated that the fusion algorithm was superior to the standard algorithm as measured by sensitivity, positive predictive value, and specificity. Use of the fusion algorithm in intensive care unit monitors would result in a greater number of true positive alarms (126 vs 118) as well as fewer false positive alarms (58 vs 223). Furthermore, the results from Table 8A demonstrated that using the fusion algorithm actually resulted in fewer missed true positive alarms than using the standard algorithm. Therefore, there was no danger that generating fewer false positive alarms with the fusion algorithm also generated fewer true positive alarms by the fusion algorithm.

The analyses of the number of alarms generated as a function of average heart rate indicated that patients with low average heart rates (ie 60-80 beats per minute) had higher numbers of low heart rate alarms. Also, patients with high average heart rates (ie 100-120 beats per minute) had greater numbers of high heart rate alarms. These findings suggest that customizing alarm limits, either manually or automatically, based on each patient's actual heart rate may result in more true positive and fewer false positive alarms in the ICU. The idea of customizing alarm settings has also been suggested in the literature.[15] Our finding that a majority (89/109, 81.7%) of the patients in this study never had their monitors' alarm thresholds changed from their default values suggests that customization of alarms limits is not often done.

The durations of a large proportion (60.2%) of the alarms generated in the study were 10 seconds or less. It is questionable whether caregivers acted upon alarms of such short duration or whether these short alarms were merely annoyances that could lead to clinicians either ignoring alarms or turning them off altogether. Further research is required to answer these questions. If alarms of short duration were determined to be unimportant to caregivers, then simply extending the delay between the time an alarm condition began and the time that an alarm actually sounds would dramatically reduce the number of false positive alarms generated by ICU monitors.

Excessive numbers of false positive alarms have been a problem for clinicians for decades. False alarms may pose a hazard for patients if doctors and nurses become desensitized to them or if audible alarms are deactivated. Finding ways to increase the true positive alarm rate and decrease the false positive rate for ICU monitors alarms should benefit both caregivers and their patients. This study has shown that using a fusion alarm algorithm to combine heart rate signals from multiple, independent patient monitors is superior to using a standard alarm algorithm that relies only on a single heart rate signal from one monitor. Compared with the standard algorithm, the fusion algorithm improves true positive alarm rates while reducing false positive alarm rates.

REFERENCES

1. Raison JCA, Beaumont JO, Russell JAG, Osborn JJ, Gerbode F. Alarms in an intensive care unit: an interim compromise. *Comput Biomed Res.* 1968;1:556-564.

2. Chambrin MC. Alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit Care.* 2001;5:184-188.

3. Lutter NO, Urankar S, Kroeber S. False alarm rates of three third-generation pulse oximeters in PACU, ICU and IABP patients. *Anesth Analg.* 2002;94(Suppl):S69-75.

4. Malviya S, Reynolds PI, Voepel-Lewis T, Siewert M, Watson D, Tait AR, Tremper K. False alarms and sensitivity of conventional pulse oximetry versus the Masimo SET technology in the pediatric postanesthesia care unit. *Anesth Analg.* 2000;90:1336-1340.

5. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med.* 1999;25:1360-1366.

6. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med.* 1994;22:981-985.

7. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med.* 1997;25:614-619.

8. McIntyre JWR. Ergonomics: anaesthetists' use of auditory alarms in the operating room. *Intl J Clin Monit Comput.* 1985;2:47-55.

9. Block FE, Nuutinen L, Ballast B. Optimization of alarms: a study on alarm limits, alarm sounds, and false alarms, intended to reduce annoyance. *J Clin Monit Comput.* 1999;15:75-83.

10. Duberman SM, Bendixen HH. Concepts of fail-safe in anesthetic practice. *Intl Anesth Clinics.* 1984;22:149-165.

11. Feldman JM, Ebrahim MH, Bar-Kana I. Robust sensor fusion improves heart rate estimation: clinical evaluation. *J Clin Monit.* 1997;13:379-384.

12. Ebrahim MH, Feldman JM, Bar-Kana I. A robust sensor fusion method for heart rate estimation. *J Clin Monit.* 1997;13:385-393.
13. Cheng H, Macaluso M. Comparison of the accuracy of two tests with a confirmatory procedure limited to positive results. *Epidemiology.* 1997;8:104-106.

14. Schatzkin A, Connor RJ, Taylor PR, Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. *Am J Epidemiol.* 1987;4:672-678.

15. Solsona JF, Altaba C, Maull E, Rodriguez L, Bosque C, Mulero A. Are auditory warnings in the intensive care unit properly adjusted? *J Adv Nurs.* 2001;35:402-406.